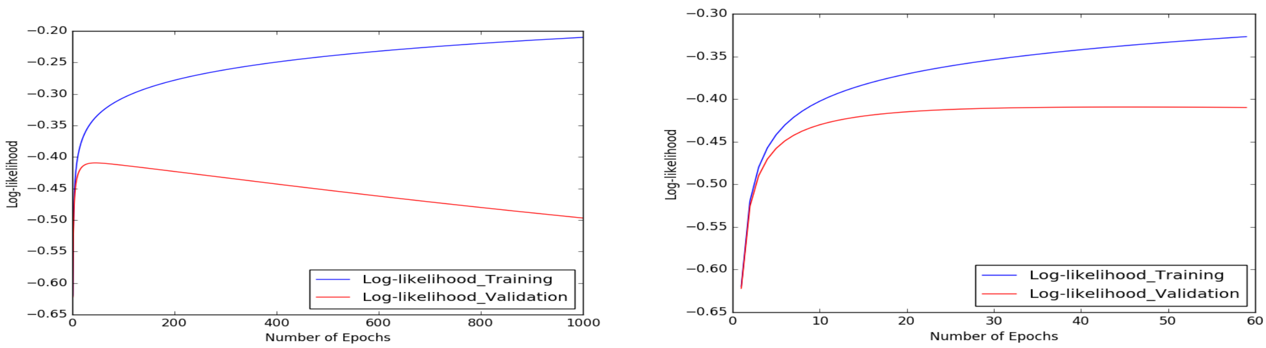
ECE521: Inference Algorithms and Machine Learning
Report of Assignment 2

Team Members:
Xi Chen (1002266013)
Yang Wu (1002495132)

Task1: Classification Error using LR

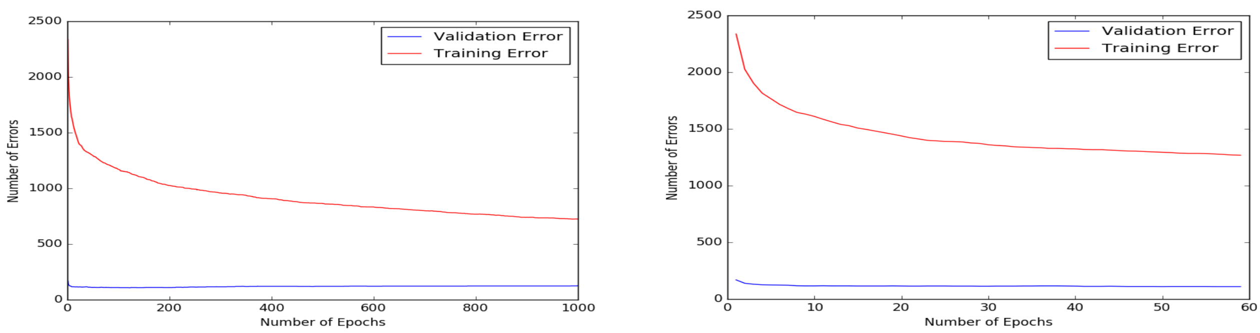
Set learning rate = 0.01 and add momentum to speed up the training with momentum rate = 0.4. We got the following figures to show: (i) the log-likelihood of the training data and the log-likelihood of the validation data vs the number of epochs, (ii) the number of training errors and validation errors vs the number of epochs.

Test Accuracy = **88.6397%**
Test Error: **309**



(a)Complete 1000 training epochs (b)Training with early stopping

Figure 1 Log-likelihood of the training data and the validation data vs the number of epochs



(a)Complete training epochs (b)Training with early stopping

Figure 2. The number of training errors and validation errors vs the number of epochs

We can see from the Figure 1 that the log-likelihood of the training data keeps rising during the training process and the number for training errors keeps decreasing as the training accuracy increases; whereas, for the validation data, the log-likelihood reaches at the maximum after **60 epochs** and starts to decrease if the training continues. Hence, we think the best time to stop the training in our model is the time when the log-likelihood of the validation data is maximum (about 60 epochs in our model). The test accuracy and the number of test error at this time are **88.64%** and **309** respectively.

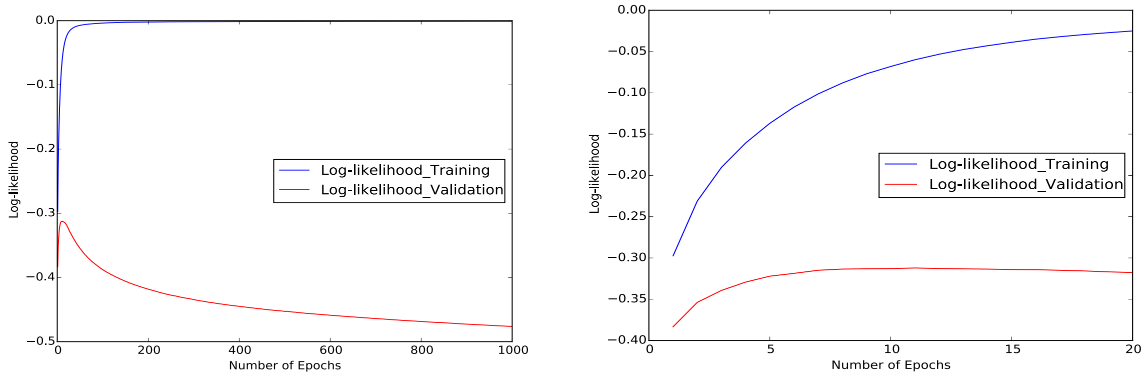
Task 2: Neural Network Training

In the neural network with one hidden layer and 1000 hidden units, we train it using different values of learning rate and got the following results:

(1) Learning rate = 0.05, training epochs = 1000

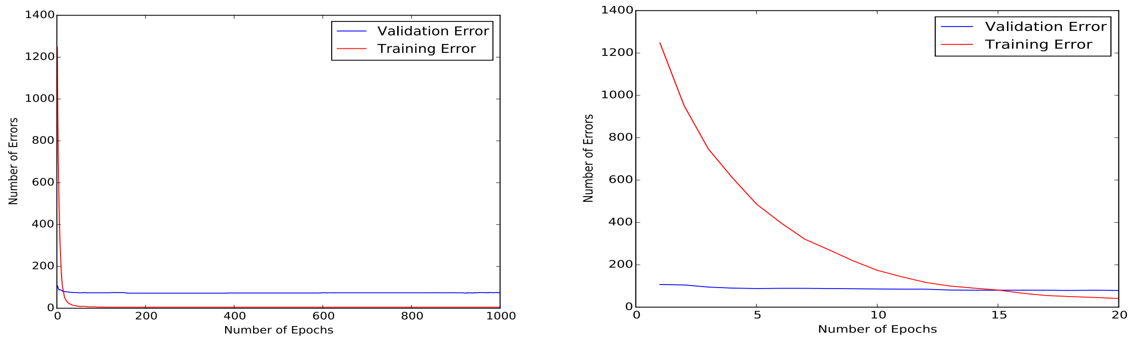
Test Accuracy = 91.0662%

Test Error: 243



(a) Complete 1000 epochs (b) early stopping

Figure 3. Log-likelihood vs the number of epochs (learning rate: 0.05)



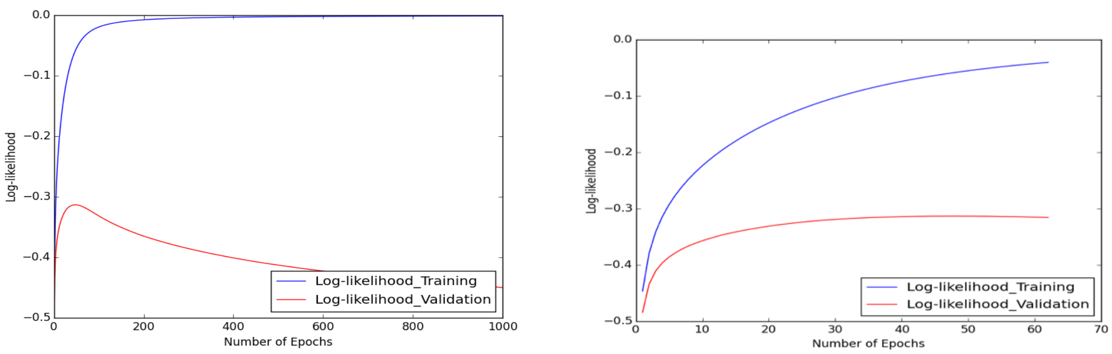
(a) Complete 1000 epochs (b) early stopping

Figure 4. The number of errors vs the number of epochs (learning rate: 0.05)

(2) Learning rate = 0.01, training epochs = 1000

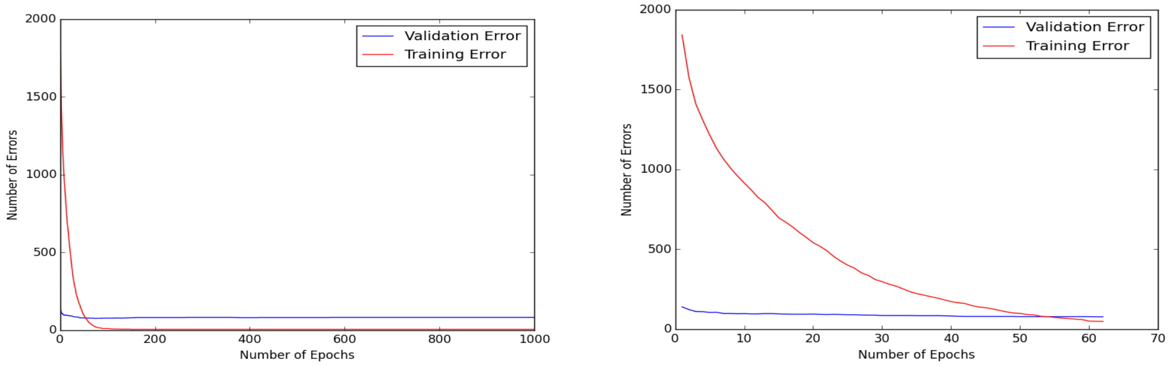
Test Accuracy = 90.7353%

Test Error: 252



(a) Complete 1000 epochs (b) early stopping

Figure 5. Log-likelihood vs the number of epochs (learning rate: 0.01)



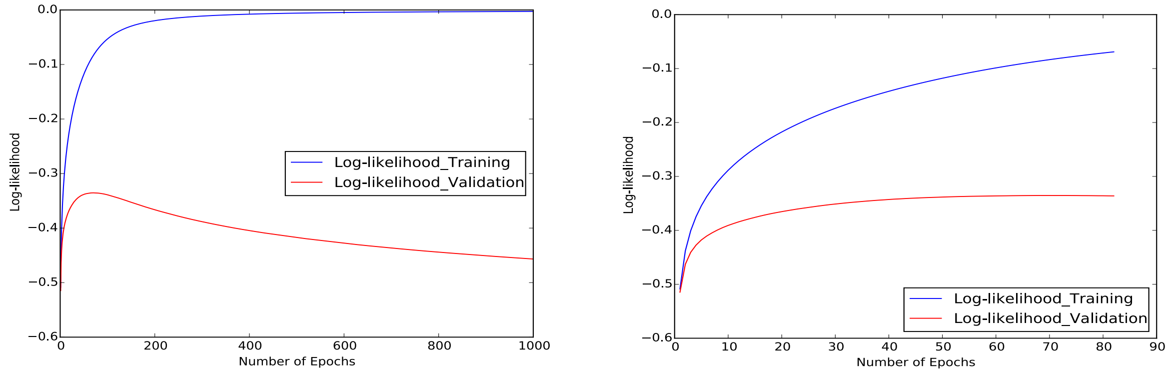
(a) Complete 1000 epochs (b) early stopping

Figure 6. The number of errors vs the number of epochs (learning rate: 0.01)

(3) Learning rate = 0.005, training epochs = 1000

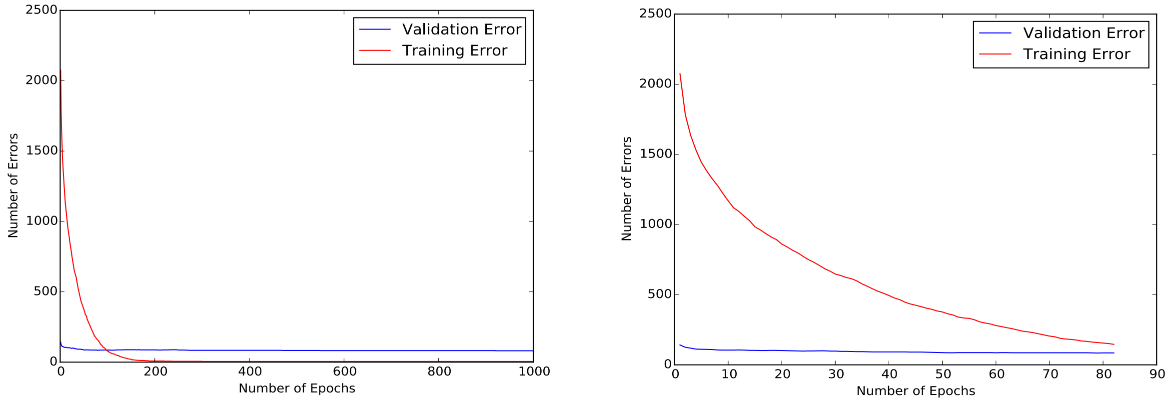
Test Accuracy = 90.4412%

Test Error: 260



(a) Complete 1000 epochs (b) early stopping

Figure 7. Log-likelihood vs the number of epochs (learning rate: 0.005)



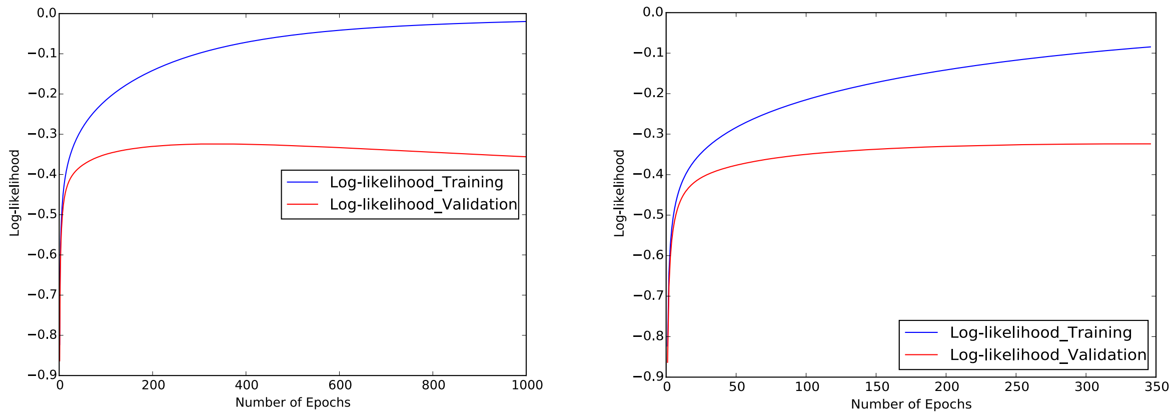
(a) Complete 1000 epochs (b) early stopping

Figure 8. The number of errors vs the number of epochs (learning rate: 0.005)

(4) Learning rate = 0.001, training epochs = 1000

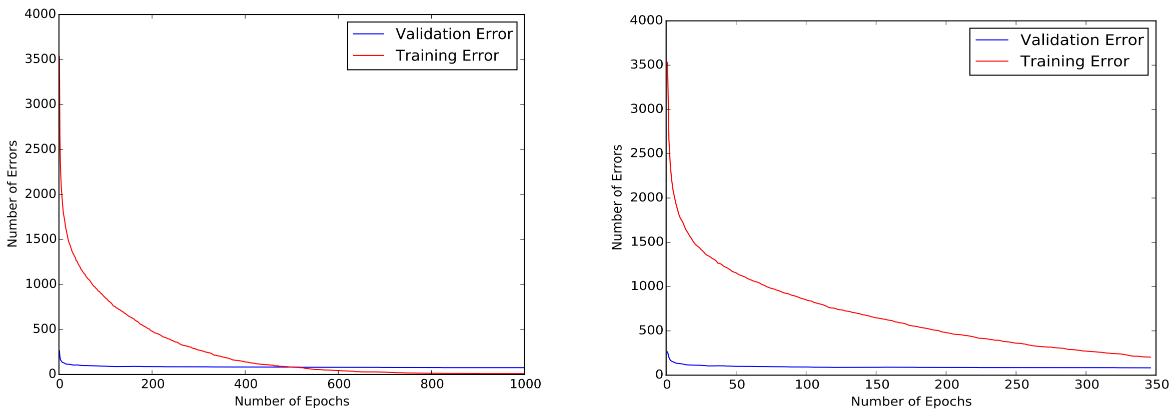
Test Accuracy = 90.1838%

Test Error: 267



(a) Complete 1000 epochs (b) early stopping

Figure 9. Log-likelihood vs the number of epochs (learning rate: 0.001)



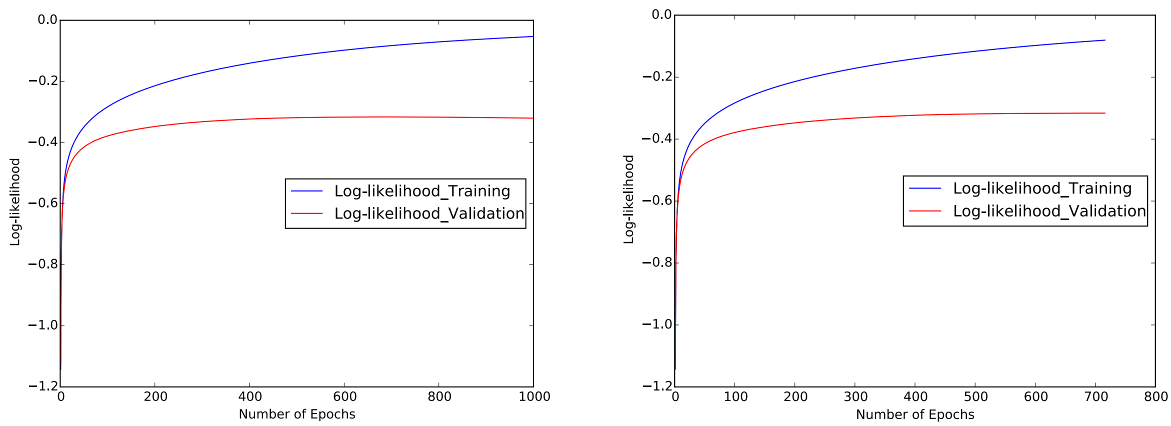
(a) Complete 1000 epochs (b) early stopping

Figure 10. The number of errors vs the number of epochs (learning rate: 0.001)

(5) Learning rate = 0.0005, training epochs = 1000

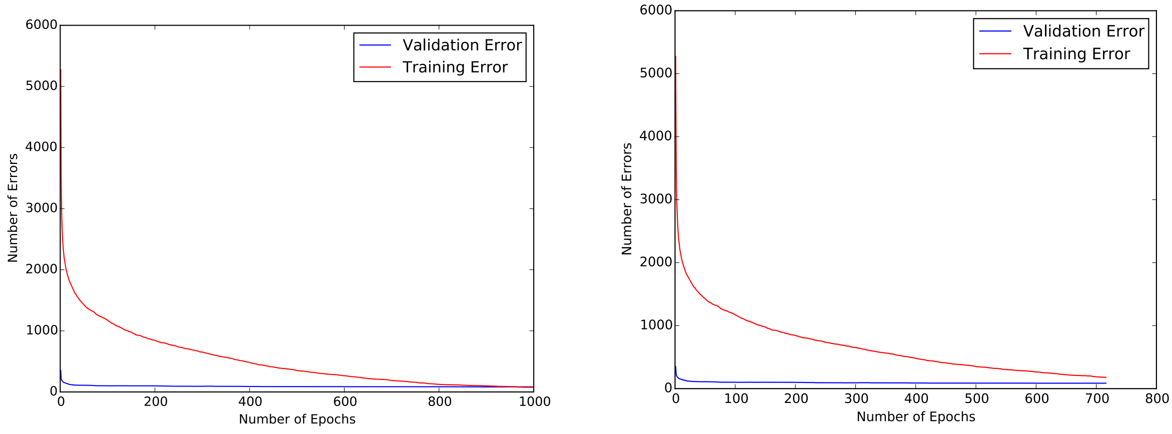
Test Accuracy = 90.7721%

Test Error: 251



(a) Complete 1000 epochs (b) early stopping

Figure 11. Log-likelihood vs the number of epochs (learning rate: 0.0005)



(a) Complete 1000 epochs (b) early stopping

Figure 12. The number of errors vs the number of epochs (learning rate: 0.0005)

The above figure 3 to figure 11 have been showing the training process of a structural identical model in addition to different learning rate (0.05, 0.01, 0.005, 0.001, and 0.0005). On one hand, as the learning rate getting smaller, the training process takes more epochs to reach optimal likelihood so that to stop itself; on the other hand, there is no direct evidence to support any kind of relation between learning rates and accuracies. With learning rate of 0.05, we get the maximum accuracy among all results, but the learning rate of 0.0005 does not give us the minima. Actually, the learning rate of 0.001 leads to a minimum accuracy.

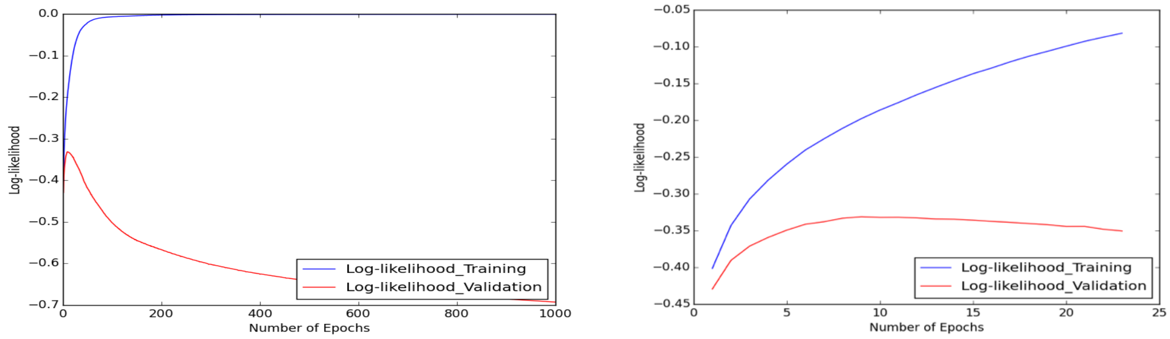
Also, after stopping at the optimal log-likelihood, the validation accuracy is not necessarily the optimal result. If we allow the training process to continue, there exists an even better result as shown in the figure 3-11 (a).

Since the task ask to select the best model using the validation log-likelihood, the model with learning rate of 0.05 is chosen here. In this model, we get **the number of test errors of 243** when we use early stopping.

<u>Task3: Number of Hidden Units</u>

We set the learning rate = 0.05 and train different neural networks with 100, 500 and 1000 hidden units, and get the following results:

(1)100 hidden units:



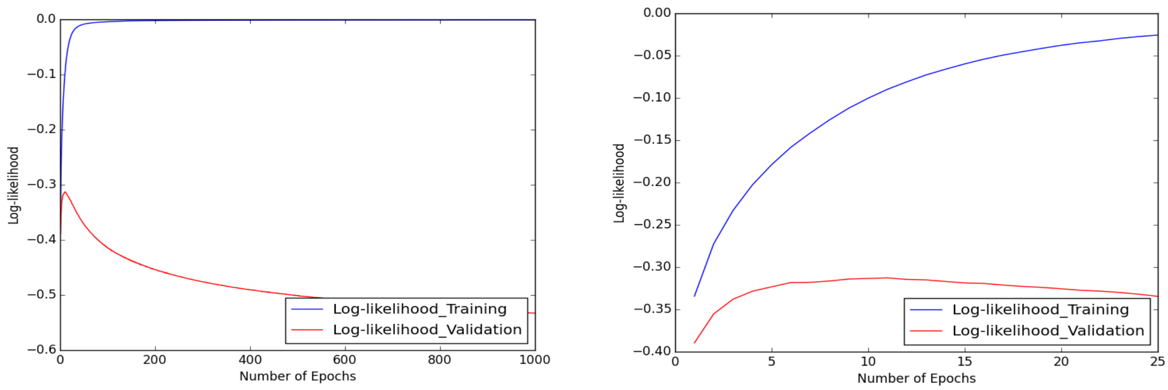
(a) Complete 1000 epochs (b) early stopping

Figure 13. Log-likelihood vs the number of epochs with 100 hidden units and 1 layer

(a) Complete 1000 epochs (b) early stopping

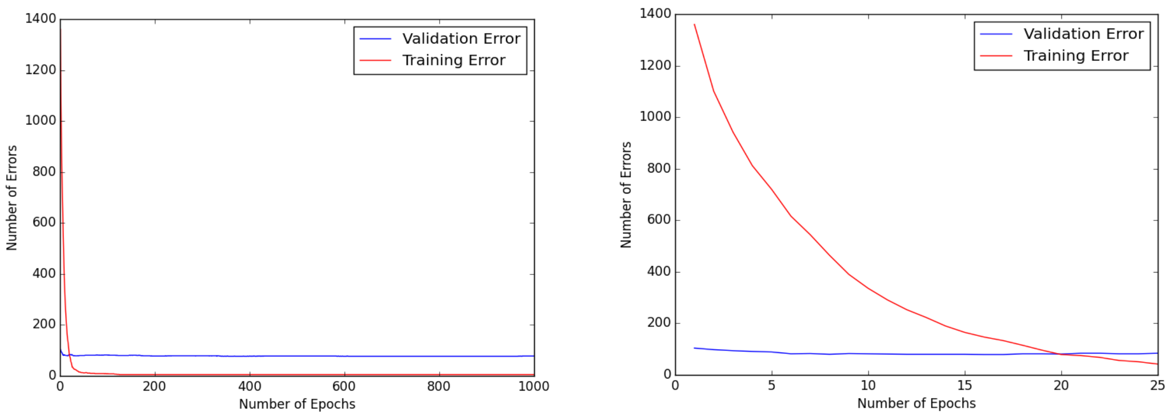
Figure 14. The number of errors vs the number of epochs with 100 hidden units and 1 layer

(2) 500 hidden units:



(a) Complete 1000 epochs (b) early stopping

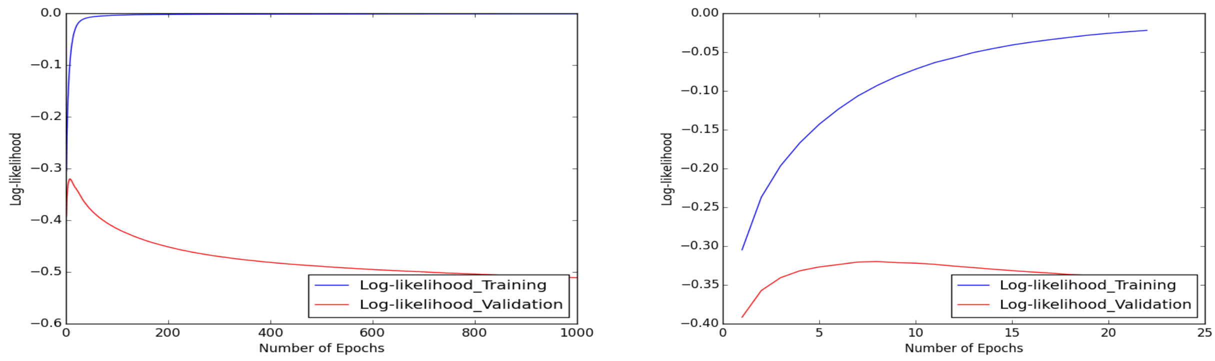
Figure 15. Log-likelihood vs the number of epochs with 500 hidden units and 1 layer



(a) Complete 1000 epochs (b) early stopping

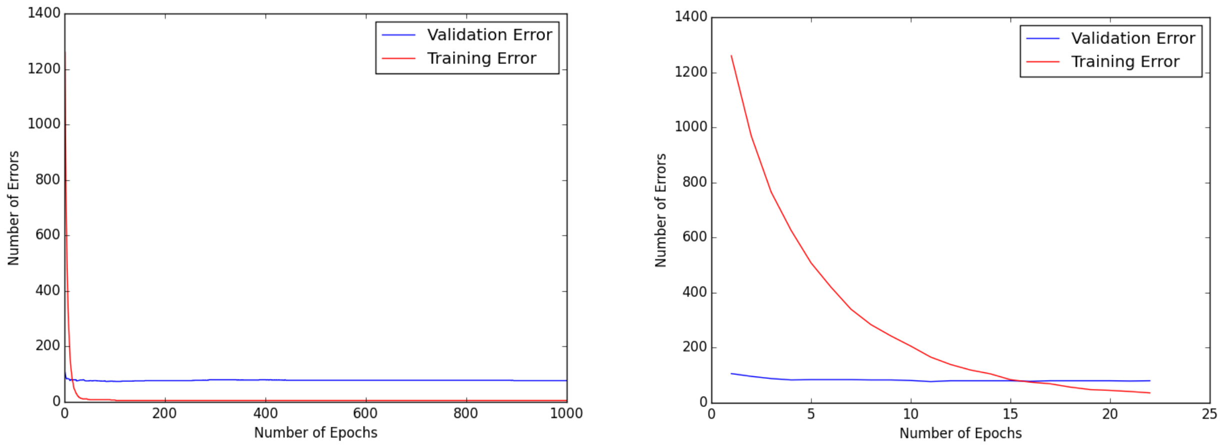
Figure 16. The number of errors vs the number of epochs with 500 hidden units and 1 layer

(3) 1000 hidden units:



(a) Complete 1000 epochs (b) early stopping

Figure 17. Log-likelihood vs the number of epochs with 1000 hidden units and 1 layer



(a) Complete 1000 epochs (b) early stopping

Figure 18. The number of errors vs the number of epochs with 1000 hidden units and 1 layer

Using the above models to classify the test cases, we got the following results:

(a) 100 hidden units:
Test Accuracy: 90.7721%
The Number of Test Errors: 251

(b) 500 hidden units:
Test Accuracy: 91.0294%
The Number of Test Errors: 244

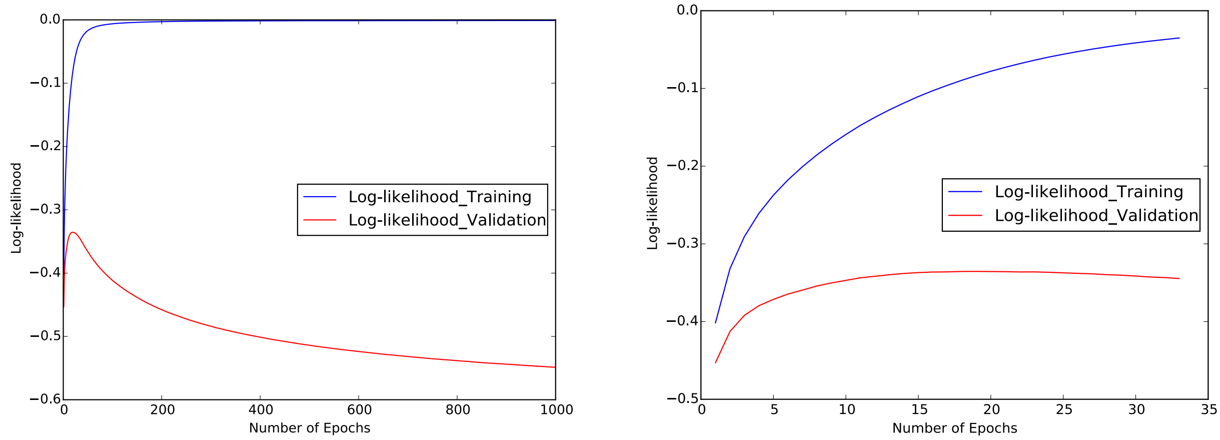
(c) 1000 hidden units:
Test Accuracy: 91.2868 %
The Number of Test Errors: 237

According to the above data, we can conclude that more hidden units can lead to a higher test accuracy and fewer test errors in our case where there are more than 15000 training samples. Sometime, there are cases with 500 hidden units having the optimal result, but the result of 1000 hidden units has higher probability of giving a better value.

Hence, in the probabilistic view, 1000 hidden units give the best result. Before a super large set of hidden units makes the model overfitting, more hidden units are more suitable for the nonlinearity of the inputs so that it can better extract features from training samples.

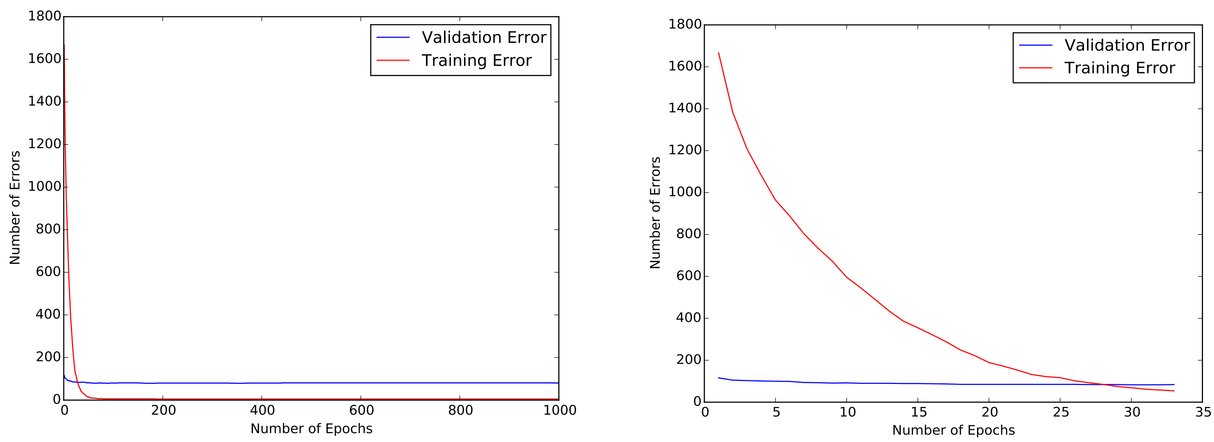
Task4: Number of layers:

In this task, we set two layers and each layer has 500 hidden units. The learning rate we use here is 0.01, the results are as follows:



(a) Complete 1000 epochs (b) early stopping

Figure 19. Log-likelihood vs the number of epochs with two layers and 500 hidden units in each layer



(a) Complete 1000 epochs (b) early stopping

Figure 20. The number of errors vs the number of epochs with two layers and 500 hidden units in each layer

Results for test cases:

Test Accuracy = 90.6618%

Test Error: 254

Comparing this value to the above data of 1000 hidden units in Task2, it seems the single layer with 1000 hidden units gives a better result. Again, the two-layer structure sometime provides better values, but it is rare in statistic. However, once if the two-layer model achieves the optimal value, rather than the usual 0.5% gain, it tends to boost a greater (1%) improvement.

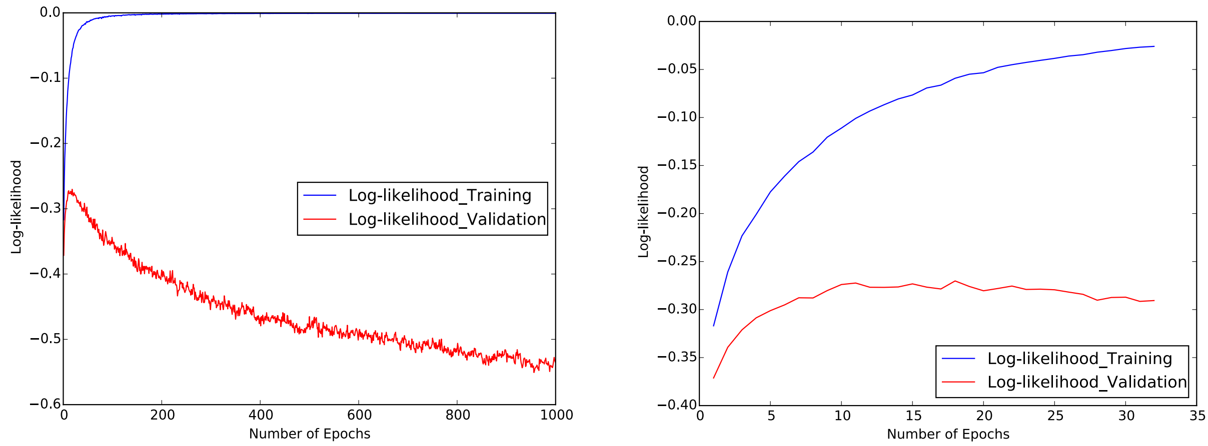
The experimental result can be explained as:

- For one-layer model, there are 784x1000x10 = 7.84×10^6 possible paths;
- For two-layer model, there are 784x500x500x10 = $7.84 \times 10^9/4$ possible paths;

Even though two models have the same number of hidden units (similar overfitting), the two-layer one with 10^3 times more possibilities leads to a smaller chance of finding the most optimal matrix.

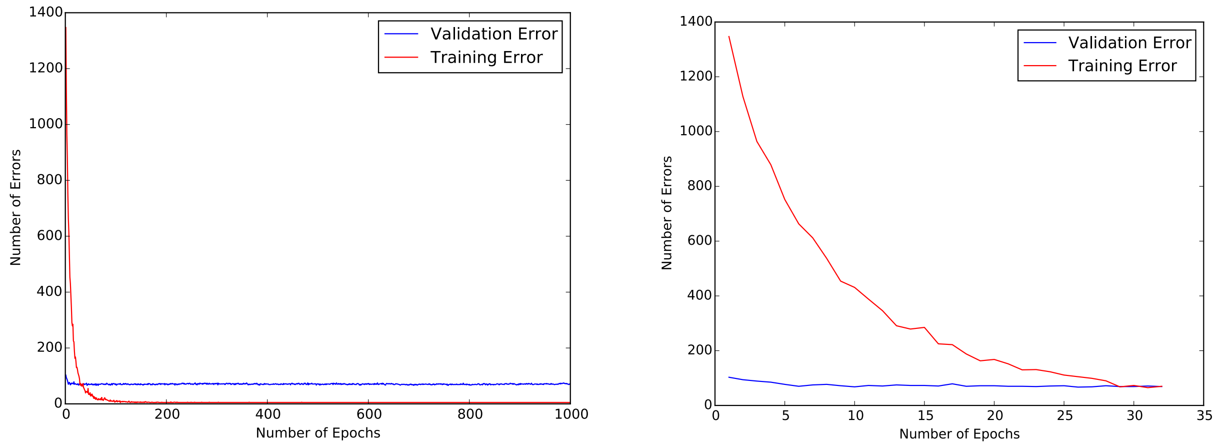
Task5: Dropout

In this task, we use the same architecture as in Task2 (learning rate: 0.05 and 1000epochs), and use dropout in the hidden layer of the neural network with dropout rate = 0.5. The log-likelihood and the number of errors vs the number of epochs are shown in the following figures.



(a) Complete 1000 epochs (b) early stopping

Figure 21. Log-likelihood vs the number of epochs with dropout



(a) Complete 1000 epochs (b) early stopping

Figure 22. The number of errors vs the number of epochs with dropout

Results for test dataset with dropout:

Test Accuracy = 91.9118%

The Number of Test Errors: 220

Results for test data set in the case without dropout (Task2):

Test Accuracy = 91.0662%

The Number of Test Errors: 243

We can see from the above figure 21 and 22 that adding dropout will result in oscillation in the errors and log-likelihood during the training process. The log-likelihood and number of errors will oscillate in a small range since every unit will be kept by the dropout rate = 0.5, thus it will help to avoid overfitting and get better performance. For the test cases, the test accuracy is higher than that in Task2, which does not use dropout, thus the number of test errors is smaller than that in Task2. (**91.9118% vs 91.0662%** and **220 vs 243**)

Task6: Crowd-Sourcing for Hyperparameter Search

In this task, we randomly pick the hyperparameter settings (learning rate, number of layers, number hidden units, dropout rate and momentum rate) and train 10 groups of data. The results of test errors and validation errors are listed in the following part:

(1) Hyperparameter settings:
Log of learning rate: -3.556016
The number of layers: 3
The number of hidden units per layer: 333, 240, 378
Dropout rate:0.5
Momentum rate:0.5

Training Results:
Test Accuracy = 90.2206%
The Number of Test Errors: 266
Validation Accuracy = 90.7000%
The Number of Validation Errors: 93

(2) Hyperparameter settings:
Log of learning rate: -2.331083
The number of layers: 2
The number of hidden units per layer: 395, 190
Dropout rate:1.0
Momentum rate:0.5

Training Results:
Test Accuracy = 90.0735%
The Number of Test Errors: 270
Validation Accuracy = 91.2000%
The Number of Validation Errors: 88

(3) Hyperparameter settings:
Log of learning rate: -2.054759
The number of layers: 2
The number of hidden units per layer: 459, 324
Dropout rate: 1.0
Momentum rate: 0.5

Training Results:
Test Accuracy = 90.9559%
The Number of Test Errors: 246
Validation Accuracy = 92.0000%
The Number of Validation Errors: 80

(4) Hyperparameter settings:
Log of learning rate: -2.479879
The number of layers: 1
The number of hidden units per layer: 208
Dropout rate: 1.0

Momentum rate: 0.3

Training Results:
Test Accuracy = 90.4412%
The Number of Test Errors: 260
Validation Accuracy = 91.3000%
The Number of Validation Errors: 87

(5) Hyperparameter settings:
Log of learning rate: -2.582159
The number of layers: 2
The number of hidden units per layer: 315, 196
Dropout rate: 1.0
Momentum rate:0.3

Training Results:
Test Accuracy = 90.2941%
The Number of Test Errors: 264
Validation Accuracy = 91.2000%
The Number of Validation Errors: 88

(6) Hyperparameter settings:
Log of learning rate: -2.769707
The number of layers: 2
The number of hidden units per layer: 438, 243
Dropout rate: 0.5
Momentum rate: 0.3

Training Results:
Test Accuracy = 90.4044%
The Number of Test Errors: 261
Validation Accuracy = 91.0000%
The Number of Validation Errors: 90

(7) Hyperparameter settings:
Log of learning rate: -2.082970
The number of layers: 3
The number of hidden units per layer: 364, 124, 232
Dropout rate:1.0
Momentum rate:0.5

Training Results:
Test Accuracy = 90.9191%
The Number of Test Errors: 247
Validation Accuracy = 91.8000%
The Number of Validation Errors: 82

(8) Hyperparameter settings:
Log of learning rate: -2.608272

The number of layers: 2
The number of hidden units per layer: 431,199
Dropout rate:0.5
Momentum rate:0.4

Training Results:
Test Accuracy = 90.8824%
The number of Test Errors: 248
Validation Accuracy = 91.7000%
The number of Validation Errors: 83

(9) Hyperparameter settings:
Log of learning rate: -3.470326
The number of layers: 3
The number of hidden units per layer: 169, 436, 360
Dropout rate:0.5
Momentum rate: 0.4

Training Results:
Test Accuracy = 89.3750%
The number of Test Errors: 289
Validation Accuracy = 90.0000%
The number of Validation Errors: 100

(10) Hyperparameter settings:
Log of learning rate: -3.219482
The number of layers: 1
The number of hidden units per layer: 247
Dropout rate:0.5
Momentum rate: 0.5

Training Results:
Test Accuracy = 89.8162%
The number of Test Errors: 277
Validation Accuracy = 90.9000%
The number of Validation Errors: 91